UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

THE NEW YORK TIMES COMPANY, NICHOLAS CONFESSORE, and GABRIEL DANCE,

Plaintiffs,

-v-

FEDERAL COMMUNICATIONS COMMISSION,

Defendant.

18 Civ. 8607 (LGS)

DECLARATION OF ERIK SCHEIBERT

I, ERIK SCHEIBERT, pursuant to 28 U.S.C. § 1746, declare the following under penalty of perjury:

1. I currently serve as the Associate Chief Information Officer for Engineering and the Chief Enterprise Architect within the Office of the Managing Director at the Federal Communications Commission ("FCC"). I have worked at the FCC, in this and other information technology capacities, since 2005.

2. In this role, I am the subject matter expert on the FCC's Electronic Comment Filing System ("ECFS"). I had this responsibility during April through June of 2017, the period covered by the FOIA request that is the subject of the Complaint filed in *The New York Times Company, et al. v. Federal Communications Commission*, 18 Civ. 8607, currently pending in the United States District Court for the Southern District of New York.

3. On or around December 19, 2017, and April 20, 2018, I participated in teleconferences in which I explained the technical features of ECFS to employees of The New York Times in an attempt to resolve The New York Times's appeal of the FCC's denial of FOIA request No. 2017-738.

4. The statements contained in this declaration are based upon my personal knowledge, upon information provided to me in my official capacity, and upon conclusions and determinations reached and made in accordance therewith.

The System Architecture of ECFS

5. ECFS is the FCC's information system that serves as the repository for public comments and other materials submitted in the course of rulemaking and other FCC docketed proceedings. ECFS allows members of the public to comment electronically on FCC proceedings and to access comments other parties have filed. A defining design feature of ECFS, therefore, is that it is an open, public-facing system. It is configured to be easily accessed and used by any member of the public interested in participating in or learning about an FCC proceeding. It is designed to accept data from the public in large volumes and in a wide variety of formats, and to manage the large spikes in traffic that the application experiences during high-profile proceedings.

6. ECFS is an application built on cloud-computing infrastructure. It consists of a number of virtual servers and other devices operating within a "virtual private cloud" separated from the public Internet by firewalls and other security measures. ECFS was built and has been operated and maintained by the FCC's Information Technology ("IT") staff.

7. A member of the public interested in using ECFS can access it in two ways from the public Internet. Using either mode of access, users can submit comments or search for and review comments submitted by others. Under the first method, a user can access ECFS through the human user interface located at the following Uniform Resource Locator ("URL"): https://www.fcc.gov/ecfs. Under the second method, a user can access ECFS through an application programming interface ("API") located at the following URL:

https://publicapi.fcc.gov/ecfs. The API allows users to extract information from and/or submit comments to ECFS in larger volumes and at faster rates than the human user interface allows.

8. ECFS gives users the option to either file comments with supporting documents ("Standard Filings") or allows users to submit their comments via an online form ("Express Filings"). For both Standard and Express filings, ECFS requires users to identify the proceeding on which they are commenting and to provide their names and postal addresses. Both the Standard and Express filing forms include the following notice: "You are filing a document into an official FCC proceeding. All information submitted, including names and addresses, will be publicly available via the web."

9. ECFS processes users' computerized requests to submit or review comments in a variety of ways.[1] Many common requests for information can be fulfilled by information stored (or "cached") external to the ECFS application by a commercial content delivery network with which the FCC contracts to help manage ECFS web traffic. Other information requests and all requests to submit comments are processed through a series of intermediate application elements (known as "hops") located internally within the ECFS application, before the requests reach the internal FCC servers in which ECFS content is stored and processed (the ECFS "application" and "database" servers).

10. These internal hops include software that distributes user requests among the computing resources available in ECFS ("load balancers"), servers that process and respond to user web requests ("web servers"), and servers that organize and filter the requests ("proxy

[1] For the purposes of this declaration, a "request" means any of the common "client" request methods defined by the hypertext transfer protocol (http). The FOIA requests at issue refer to "GET" and "POST" requests. A successful GET request will retrieve data from ECFS, while a successful POST request will submit new data into ECFS.

servers"). To allow these application elements to communicate with each other and to process user requests, each of these internal ECFS servers is assigned a private Internet Protocol ("IP") address. Processing a single request through multiple hops will take a certain period of time, usually less than a second, but sometimes longer in high traffic periods.

11. The FCC has a strong security interest in protecting ECFS's internal configuration from public disclosure. In particular, publicly disclosing the internal IP addresses which the FCC uses to link various ECFS components will give attackers a roadmap through the system, and threaten the confidentiality, integrity or availability of the information in the ECFS database and the system itself.[2] Aided with knowledge of these internal IP addresses, attackers can access the internal system elements that manage system traffic and process user requests. In addition, attackers can potentially delete or alter information in the system, or make the system publicly inaccessible. More generally, if internal information about the ECFS system is publicly disclosed, attackers will have valuable information about how the FCC secures and configures its other information systems, including systems operated by the FCC's Enforcement Bureau, its Public Safety and Homeland Security Bureau, and its Office of Inspector General.

The June 2017 FOIA Request

12. On June 22, 2017, plaintiff Nicholas Confessore, on behalf of The New York Times, filed a FOIA request ("June 2017 Request") with the FCC. The request was assigned FOIA Control No. FCC-2017-000764 (generally, the "FOIA Request"). The June 2017 Request stated:

[2] The Federal Information Security Modernization Act, the law under which the FCC and other Federal Government agencies operate their information systems, defines confidentiality, integrity, and availability as the three objectives of "information security." 44 U.S.C. § 3552(3).

> Please provide the web server logs for comments submitted for Federal Communications Commission docket No. 17-108 between 4/26/17 and 6/7/2017. I would like the logs for requests submitted via both to https://www.fcc.gov/ecfs/filings/ and any submissions through the FCC's API (application programming interface). For each comment, please include the following information: 1) Server logs for both GET and POST requests 2) The date/time stamp of each request 3) The full query including query strings 4) The IP address of the client making the request 5) The browser USERAGENT 6) The following headers when available: Accept, Accept-Encoding, Accept-Language, Connection, Host, DNT, Upgrade-Insecure-Requests, Via, X-Forwarded-For.

A true and correct copy of The New York Times's June 2017 Request is annexed hereto as **Exhibit A.**

13. By letter dated July 21, 2017, the FCC denied the June 2017 Request, stating that the information requested may be withheld in full under FOIA Exemption 6 because it "includes personably identifiable information and therefore cannot be released." A true and correct copy of the FCC's July 21, 2017 denial letter is annexed hereto as **Exhibit B.**

14. By letter dated July 25, 2017, The New York Times appealed the agency's denial. A true and correct copy of The New York Times's appeal of the FCC's denial is annexed hereto as **Exhibit C**.

15. Given the system architecture and the security concerns described above, Mr. Confessore's June 2017 Request for "web server logs for comments submitted for FCC docket No. 17-108" was not a simple, straightforward request. The ECFS application does not log comments in the way Mr. Confessore assumed it did for the purpose of the June 2017 Request.

16. Server logs are files that a server automatically generates when it or another system element performs its activities. The data in server logs are not "submitted" by users; they are artifacts of the machine-to-machine communication necessary to execute a request. The FCC's server logs also help the agency detect patterns that indicate potential malicious software ("malware") or other security threats. ECFS does not have a single set of server logs that

comprehensively documents ECFS's activities. Instead, multiple server logs document the activities of the various internal servers and other elements involved in the operations of ECFS. Significantly, ECFS does not assign a unique identifier to incoming requests in order to accurately track the processing of user requests across multiple server log entries. If ECFS had such a feature, FCC IT staff could reliably map the processing of a user request throughout the servers, or hops, in the ECFS system. Because ECFS lacks such a feature, however, it is impossible for FCC to directly and definitively track a public comment in the database back to "the IP address of the client making the request," as Mr. Confessore requested. The data captured in internal server logs typically records only the immediately preceding internal hop, rather than the entire path of the request.

17. To <u>approximately</u> track a single comment back to the original user request, FCC would have to engage in a painstaking process of working backwards from the date and time the comment appeared in the ECFS database, as opposed to relying on a reliable unique identifier that would attach to and track the request as it progressed through multiple internal hops. The only way FCC can approximately track a single comment is to laboriously retrace the request's path through the multiple internal hops, as those hops are recorded in multiple server logs, back out to the original request made from the public Internet. The retracing process would allow FCC to identify several requests made close in time to the second the comment appears in the database, and guess which one is the actual originating request. However, FCC cannot directly and conclusively correlate one ECFS request with one ECFS comment.

18. Several other factors complicate the process for tracking a single comment back to the original user request. In busy periods, for example, ECFS receives thousands of requests a minute, so there can be a large number of requests that correlate very closely in time with a

specific comment. Time correlation is also complicated by the fact that the individual servers' timestamp mechanisms are not fully synchronized. Furthermore, the server logs record user requests for all ECFS dockets, so a POST log entry will not disclose whether the user was posting a comment in FCC Docket No. 17-108 or in another of the hundreds of proceedings that were active in April and May 2017.

19. Separately, The New York Times's request also created security concerns. If the FCC provides the server logs necessary to trace a submitted comment back to the IP address of the requester, it will necessarily disclose the IP addresses of the internal system elements (the hops) through which the request traveled. As explained above in Paragraph 11, this will compromise the security of the FCC's computer systems.

20. The New York Times and the FCC subsequently engaged in email and telephone communications in an attempt to reach a consensual resolution regarding the FOIA Request. I personally explained the above-described technical limitations and security concerns to Mr. Confessore and his counsel during two telephone conference calls I participated in on or around December 19, 2017, and April 20, 2018.

The New York Times's Amended Requests
Dated September 22, 2017 and December 21, 2017

21. The New York Times amended its FOIA Request through e-mail correspondence on September 22, 2017, and a letter on December 21, 2017. The September 22 amendment eliminated certain types of header fields from the original request. The December 21 amendment further narrowed the requested information to just the comment itself, the originating IP address, the date and time stamp, and the User-Agent header. True and correct copies of those communications are annexed hereto as **Exhibit D and Exhibit E**, respectively. These amendments reduced the number of log elements The New York Times sought in its original

FOIA Request. Specifically, the September 22 modification eliminated the following header information from the original FOIA Request: Accept, Accept-Encoding, Accept-Language, Connection, Host, DNT, Upgrade-Unsecure-Requests, Via. *See* Exhibit D. The December 21 amendment further eliminated the X-Forwarded-for header and the full query, including query strings, from the original FOIA Request. *See* Exhibit E. These modifications, however, did not change Mr. Confessore's request to correlate a submitted comment with a specific ECFS request. As described above, the configuration of the ECFS system elements makes this type of correlation highly complicated and burdensome. Therefore, these "narrowed requests" did not reduce the technical difficulty of correlating a comment to its originating ECFS request. In addition, the amended requests still sought the intermediate logs necessary to trace a submitted comment back to the IP address of the requester. Thus, the requests still raised the security concerns discussed above by revealing sensitive information about the FCC's network architecture.

The New York Times's Amended Requests
Dated May 7, 2018 and August 31, 2018

22. By letter dated January 29, 2018, the FCC submitted a supplemental response to its July 2017 denial of The New York Times's June 2017 Request. A true and correct copy of the letter is attached as **Exhibit F**. In that letter, FCC stated that the records sought by the June 2017 Request are exempt from disclosure because IP addresses are protected by FOIA Exemption 6. FCC also stated that the requested server logs are subject to withholding under FOIA Exemption 7(E) because revealing the logs would reveal "information about how the Commission protects the security of the ECFS and its other information assets."

23. By letter dated February 26, 2018, The New York Times appealed the FCC's supplemental denial. A true and correct copy of that letter is attached as **Exhibit G.**

24. By letter dated May 7, 2018, The New York Times amended Mr. Confessore's request again. A true and correct copy of the letter is attached as **Exhibit H**. In this May 7, 2018 letter, Christina Koningisor of The New York Times's Legal Department expressed The New York Times' understanding of the configuration of the ECFS server logs, and the obstacle this created to the FOIA Request. This letter stated:

> The FCC can't definitively link [the user IP address] to [the User-Agent header] because i) there's no unique identifier recorded in each set of logs, and ii) it takes time to flow through each layer, so the timestamps for a single comment might be slightly different between the two sets of logs.

See Exhibit H. The New York Times also acknowledged the FCC's security concerns about disclosing these internal IP addresses in its May 7, 2018 letter and agreed that the FCC could withhold these addresses. *See id.* at 2.

25. In its May 7, 2018 letter, The New York Times narrowed its prior FOIA request to seek:

> [L]ogs from the FCC's web servers handling requests to www.fcc.gov/ecfs/filings and the FCC's API between April 26, 2017 and June 7, 2017, with any non-originating IP addresses removed using a method [of replacing each non-originating IP address with a unique identifier], but retaining any User-Agent headers and originating IP addresses, along with their respective timestamps. Additionally we're requesting the comments, names and timestamps in ECFS submitted between the same dates.

Id. at 3 (hereinafter "May 2018 Request").

26. The May 2018 Request sought at least two log files: one log showing the originating IP addresses and another log showing "User-Agent" information of ECFS requests made between April 26, 2017, to June 7, 2017. *Id.* at 1. Specifically, the amended request dropped the requirement that the FCC attempt to directly correlate a submitter's IP address with the submitter's comments. Instead, to address the concern that revealing the IP addresses of our internal servers will pose a security risk, The New York Times asked the FCC to employ a "find-

and-replace" process to replace these internal IP addresses with newly-created identifiers which would show that the two log files were associated with each other. *Id.* at 2. For example, if a server's IP address is "100.1.1.1," or "100.1.1.2," under The New York Times's proposal, FCC will be required to find and replace that address with the unique identifier "A-1," or "A-2," respectively. *Id.* at 2–3.

27. In its May 2018 Request, The New York Times explained that it will use this information to perform a "statistical analysis of the data." *Id.* at 2. I understand this to mean that The New York Times will attempt to correlate a submitter's IP address and User-Agent information to the submitter's comments in FCC Docket No. 17-108. The "User-Agent" field contains specific information about a user's computer system, such as the operating system, operating system version, browser version, the browser platform, and the user's language settings. Because the User-Agent field contains specific information relating to a user's computer system, it will help someone identify particular users and distinguish between two users who share the same IP address. That information could then potentially be matched with users' corresponding comments in ECFS, which contain the commenters' provided names and addresses, allowing The New York Times to link the information contained in the server logs to a particular person.

28. By letter dated August 31, 3018, The New York Times modified the FOIA Request again ("August 2018 Request"). A true and correct copy of that letter is annexed hereto as **Exhibit I**. The August 2018 Request sought:

> [L]ogs from the FCC's web servers handling requests for docket no. 17-108 to www.fcc.gov/ecfs/filings/ and the FCC's API between April 26, 2017 and June 7, 2017, with any non-originating IP addresses removed using a method like the one described in the May 7, 2018 letter, but retaining any User-Agent headers and originating IP addresses along with their respective timestamps.

Id. at 1. Other than eliminating that portion of the May 2018 Request which sought "comments, names, and timestamps in ECFS for comments submitted between the specified dates," *id.*, the August 2018 Request was identical to the May 2018 Request. As discussed above, the "comments, names, and timestamps in ECFS for comments submitted between the specified dates" are already publicly available.

29. The May 2018 and August 2018 Requests appear aimed to correlate different logs in order to associate a user's IP address with the user's User-Agent header, but it is not necessary to correlate different logs to find this data. In ECFS, one server log, referred to as the "API proxy server log," displays both the user's IP address and User-Agent header. However, this type of server log is not limited to requests associated with a particular FCC proceeding. Nor does this server log identify the FCC proceeding to which an ECFS request is ultimately directed. It also contains other types of ECFS requests besides requests to post comments, such as requests to download comments. Therefore, limiting this API proxy server log to only posts related to a particular proceeding like FCC Docket No. 17-108 with a high degree of confidence in the resulting data is not possible for the reasons described above at Paragraphs 16–18.

30. If the FCC staff were required to approximately identify the entries in the API Proxy server log that are associated with a particular docket, such as FCC Docket No. 17-108, we would need to analyze the logs and comments to create and test a script especially for that purpose. To be clear, this is unlike querying a relational database. Unlike a relational database, the script will need to match records based on "most likely" matches, and will require the script to analyze many records for each potential match. Unlike a database, it is not possible for us to match records based on common fields or by simply running a search or sorting the two data sources by time and assuming that the records will line up. Instead, IT staff would need to

analyze the API proxy server logs and the comments, find ways to identify "most likely" matches based on variable time stamps, write and test a script to correlate the comments to the API proxy server logs, and then validate the results. Moreover, it is not possible to correlate the comments and API proxy server logs with a high degree of statistical confidence. For example, for a given time span of comments, for example, those made within a single second, we would need to make a judgment call about which proxy log entries best match comments, based on the time gap between the log and comments, and then filter out the log entries matched to comments to proceedings other than FCC Docket No. 17-108. The time gap varies over time, and therefore makes a high confidence match (or even reasonable match) difficult to achieve. It is possible to include additional log entries from servers located between the API proxy servers and the database to improve the confidence of the matching, but this significantly complicates the matching effort because there are three data sources to match up. Again, with either method, the result would be a highly imperfect estimation of which IP addresses and User-Agent string pairs are associated with comments filed with FCC Docket No. 17-108. There is no one "right" way to carry out this process—it would require the IT staff to exercise our judgment about what degree of statistical confidence is optimal and how best to achieve that end.

31. Assuming that The New York Times wants all the server logs for ECFS as their May 2018 and August 2018 Requests state, and not just the API proxy server logs described in Paragraph 29 above, fulfilling such a request involves far more than the simple "find-and-replace" process described in The New York Times's May 2018 Request and incorporated by reference in its August 2018 Request. The FCC cannot simply run a search on a database. Instead, these amended requests require the FCC to use significant IT resources to create a new record that neither exists nor is maintained as part of its operations.

32. To create the new records sought by the May 2018 and August 2018 Requests, FCC technical specialists will first have to design and write another script that is capable of analyzing each element of the data rows contained in the responsive server logs, and then extract and modify the requested elements (*e.g.*, the internal IP addresses) from these logs. There are significant technical hurdles to developing such a script.

33. For example, due to the cloud-based architecture of ECFS, in which IP addresses are dynamically assigned to different servers as supply and demand for the service changes, the script will need to be tailored to properly replace *each* version of that server's IP address with a unique identifier. Once written, the script will then need to be tested and manually checked to ensure that it operates as intended. The script will also need to be validated using multiple data sets to ensure it will properly scale if applied to the entire set of server logs. The process of developing and testing the script to create the requested record will likely take over a week of staff time. To create the actual document The New York Times is requesting, the FCC will then have to devote significant computing resources to run the script to the tens of millions of lines of server log rows that exist for the requested time period.

IP Addresses and User-Agent Headers Can Be Linked to Individual Persons

34. The New York Times's May 2018 and August 2018 Requests also raise separate issues regarding the privacy of ECFS users. These requests seek the IP addresses of users who made requests to ECFS during the specified period, as well as the logged User-Agent headers, which indicate which Web browser and operating system each user employed to access ECFS. While ECFS makes public the names and postal addresses that commenters provide with their submissions (and provides notice to commenters that it is doing so), it does not disclose the IP addresses or other technical information recorded in the FCC's logs as users interact with the

system. The FCC has consistently denied this part of Mr. Confessore's requests on privacy grounds, because the requested information can be used to identify and potentially harm particular individuals.

35. An IP address is a unique string of numbers that identifies each device on the Internet. Computers, servers, networked printers, and other networked devices are each assigned an IP address, allowing IP traffic to be properly routed to and from that device. An IP address may be static, meaning that it remains the same over an extended period of time. An IP address may also be dynamic, meaning that it is subject to change. While IP addresses are associated with devices, they often can be reliably linked to individual persons. Through a process sometimes called "identity resolution," digital advertisers combine IP addresses, browser information, and other digital identifiers with more traditional off-line identification information (such as names, addresses, etc.) to create detailed profiles of individual consumers. For example, The New York Times's current Privacy Policy describes how it combines "device information" such as IP addresses, geolocation, browser type, and browser language with other personal information about its users. *See* https://help.nytimes.com/hc/en-us/articles/115014892108-Privacy-policy/, paragraph 4(a).

36. In its May 2018 Request, The New York Times states that it intends to use the requested log information to perform a "statistical analysis," in which it will do the work of matching IP addresses—which are stored in the FCC's server logs and are not publicly available—with individual commenters who made submissions to ECFS during the requested period, by looking up the public comments posted to FCC Docket No. 17-108. *See* Exhibit H, at 2. As discussed above, the User-Agent and time stamp information sought by The New York Times will also help it identify individual users. That is, it will allow The New York Times or

another FOIA requester to link a commenter's name and address to his or her IP address. Once The New York Times, or any member of the public who obtains the same FOIA records, establishes this link, the IP address will serve as an individual identifier, not just for the purposes of ECFS, but also in other contexts. Data sets that combine consumers' "online" and "offline" personal information have a number of applications in the commercial digital marketplace, such as identity verification, fraud detection, and marketing. As discussed in Paragraph 8 above, the FCC does not give ECFS users any notice that their log data can be used for such purposes.

37. The privacy risk to ECFS users is not limited to commercial exploitation of their personal information. Malicious actors (motivated by their opposition to a commenter's views, financial gain, or some other motive) could use this information to commit identity theft or otherwise harm the user. For example, the details disclosed in the User-Agent header information about a given user's system can potentially inform malicious actors as to whether the user is employing an outdated browser or an operating system with a vulnerability. By knowing this system information and the associated system's IP address, a malicious actor can potentially exploit that vulnerability.

I declare under penalty of perjury that the foregoing is true and correct.

Dated: 3/14/19, 2019
Washington, D.C.

ERIK SCHEIBERT